Mineração de Dados em Biologia Molecular

Sobre o curso

André C. P. L. F. de Carvalho
Monitora: Valéria Carvalho

Objetivo do Curso

Fornecer ao aluno os principais conceitos de Mineração de Dados, enfatizando as técnicas mais utilizadas para aplicações práticas em Biologia Molecular

Tópicos do Curso

- Descoberta de conhecimento em conj. de dados
- Data Warehouse e OLAP
- Análise e preparação dos dados
- Pré-processamento dos dados
- Mineração de dados
  - Preditiva
  - Descritiva
- Ferramentas
- Aplicações em Biologia Molecular

Final do curso

- Entender os principais passos da descoberta de conhecimento de conjuntos de dados biológicos
- Ser capaz de sumarizar estatisticamente um conjunto de dados biológicos
- Compreender o processo de preparação dos dados
- Ser capaz de produzir e analisar modelos utilizando técnicas de mineração de dados
- Ler e compreender artigos científicos sobre mineração de dados em biologia molecular

Exercícios

- Por em prática o que for visto durante o curso
  - Preparação de dados
  - Implementação
  - Realização de experimentos
  - Análise de resultados
  - Bem escrito

Material didático

- Conteúdo dos capítulos cobertos nos livros indicados
  - Ou outros livros de MD, que cubram os tópicos visto
- Os slides do curso têm tudo, menos o essencial

Projeto

- Utilizar MD para resolver problema real
  - Dados públicos
  - Detalhes a serem definidos depois

Etiqueta de aulas

- Chegar no horário da aula
- Pedir licença para entrar e sair da sala
- Usar palavras "mágicas" por favor, com licença, obrigado (a) e desculpe
- Não conversar durante a aula
- Levantar o braço para fazer perguntas e comentários
- Não ler outro material durante a aula
- Desligar celular durante a aula
- Colocar lixo no lixo
- Não copiar de colega ou site material a ser avaliado

Avaliação

- Provas:
  - Duas provas normais
  - Provas curtas, uma após cada duas aulas
- Chamada oral
  - Alunos podem ser perguntados em aula, valendo nota
- Trabalhos:
  - Um por laboratório e um projeto final

Datas

- Prova 1: 11/10
- Prova 2: 22/11
- Projeto: 6/12

Cálculo da Média

- Sejam
  - MC = Média Chamada
  - MP = Média Aritmética das Provas
  - NT = Nota Trabalho
  - MF = Média Final
  - Se $MP \geq 5$ e $NT \geq 5$ $\rightarrow$ MF = (4MP + 3MC + 3MT) /10 ou (6MP + 4NT) / 10
  - Se $MP < 5$ ou $NT < 5$ $\rightarrow$ MF = menor valor entre MP, MC e NT

Recuperação

- Só terão direito à recuperação os alunos com **3.0** $\leq$ **MF** $\leq$ **5.0** e freqüência superior a 70%
- Observação:
  - Será dada a nota 0.0 (zero) para cópia parcial de programa ou prova, sendo o problema levado para a coordenação do curso

Livros para o Curso

- K. Faceli, A. Lorena, J. Gama, J. e A. de Carvalho: Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina, Editora LTC, 2011
- I. H. Witten e E. Frank, Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. Morgan Kauffman, 2011 (terceira edição)
- J. Han e M. Kamber, Data Mining: Concepts and Techniques, 2nd edition, Morgan Kaufmann, 2006
- V. Kumar, M. Steinbache P. Tan, Introduction to Data Mining, Pearson / Addison-Wesley, 2006

Pacotes e Conjuntos de Dados

- WEKA
  - http://directory.google.com/Top/Computers/Artificial_Intelligence/Machine_Learning
- R
  - http://lancet.mit.edu/ga
- Machine Learning Data Repository UC Irvine
  - http://www.ics.uci.edu/~mlearn/ML/Repository.html

Perguntas